ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 8, August 2013

A Relative Structure Similarity Method For Multiple Ontologies Alignment

Ali I. El-Desouky[1], Hesham A. Ali[2], Rabab S. Rashed[3]

Computers and Systems Department, Faculty of Engineering, Mansoura University, Egypt[1]

Computers and Systems Department, Faculty of Engineering, Mansoura University, Egypt[2]

Electrical Engineering Department, Faculty of Engineering, Kafr Elsheikh University, Egypt[3]

Abstract: Knowledge in domain is expressed with the help of ontology which is scattered all over its space. Using ontology gives a share in increasing precision. Different ontologies may represent the same domain, thus includes different terms that equivalently refer to the same meaning and vice versa. This results in different structures for ontologies. That's why it is necessary to relate concepts and keywords within the same domain. One of the efficient ways to relating domain knowledge representation is ontology alignment and mapping. The main objective of ontology mapping is to determine the relationship between concepts and find the semantic mapping between ontologies. This problem lies at the heart of numerous information processing applications. As the same domain knowledge are described by different ontologies differ in modeling or structure or language which leads to heterogeneity.
To overcome this problem, this paper uses different similarities to find the relationships between ontologies. String similarity but it only considers the same term that represents different meaning. So semantic similarity is needed to be employed for higher efficiency, so semantic similarity using WorldNet database is explored and the results of pervious similarities are considered the primitive similarity to overcome all heterogeneity problems by structure similarity based on common subsume concepts and semantic similarity by IC(information content) and modify calculation of IC to consider all concept attributes. Structure similarity and semantic similarity are combined and called relative similarity.
This paper suggests a modified structural similarity method called a relative structure similarity that present a way for solving the heterogeneity between ontologies based on entities that have a common subsumed which have a relationship and information content(IC). The proposed method presents a scheme based on the real relationships across ontologies, and modification of calculation of information content using intrinsic information of ontologies to overcome of drawbacks of the methods by taking into account entity attributes. Comparison studies are made to test the validity of the proposed method.

Keywords: ontology Alignment, mapping, structure similarity, Information content(IC), Lowest Common Subsumer (LCS).

I. INTRODUCTION

Nowadays there are a lot of distributed web pages connected to cover the human requirement, but they are only human understandable. But for better processing of the content of those pages' content, they need to be understandable by machines and software agents. This vision led to a next generation of the web namely, what's known as semantic web [1]. Semantic Web is not only to present the information, but for computers to read and process the information in the web pages, and extract knowledge from this information. The computer can understand the information in the Semantic Web using a data structure called ontology. Ontology provides a knowledge representation in a particular domain; it defines concepts (classes and properties) in a given domain, and shows the relationships between the defined concepts [2].
Different ontologies may be developed to describe a particular domain, so they may use different terms, data formats, modeling language and structures to represent certain knowledge [3]. Ontology establish a common vocabulary for community organization to communicate with each other, it is difficult to build standard ontology to cover all requirement for all purpose and applications, so many ontologies are built to solve this problem which leads to heterogeneity problem of these ontologies[4]. The heterogeneity between different ontologies may be in languages, vocabularies, or modeling for the same vocabularies. This problem can be solved by building a standard ontology and a standard knowledge representation, or by drawing relations between knowledge sources (ontologies).
Ontology matching is the process of finding the relations between ontologies, while alignment is the result of matching process expressing declaratively these relations. Ontology mapping refers to an identification of identical concepts or relations between different ontologies. Ontology mapping is a fragment of alignment task [5]. The matching operation determines the alignment for a pair of ontologies. Ontology matching consists of generating an alignment from two (or more) ontologies. Figure1 depicts relationship between matching and alignment, where matching process between elements of two ontologies using external resource (WorldNet), and

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 8, August 2013

description parameters of elements is the essential step to obtain alignment between ontologies. Sometimes mapping and alignment are used interchangeably [6]. Ontology matching is an important issue in any application that communicates through ontologies such as semantic web browsing, catalogs integration, ontology evaluation, multi-agent communication and query answer, where in semantic web browsing uses matching for dynamically annotating web pages with partially overlapping ontologies [5]. Catalog integration uses matching for offering an integrated access to on-line catalogs, ontology evolution uses matching for finding the changes that have occurred between two ontology versions, multi-agent communication uses matching for finding the relations between the ontologies used by two agents and translating the messages they exchange [7,8,9],and query answering uses ontology matching for translating user queries about the web.

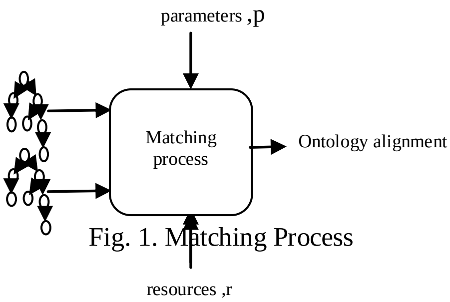


Fig. 1. Matching Process

Mapping and alignment between two ontologies is one to one function between ontology elements (concepts and properties), by comparing the similarity between a pair of elements. Ontology mapping doesn't modify ontology, the output of mapping is a pair of ontology elements with their computed similarity [10]. Any ontology mapping process is based on the following steps:

1-feature engineering: transforms ontology into internal representation

2-search step selection: select a pair of elements from ontologies based on ontology mapping, the set of pairs constitutes the search space of the method.

3-similarity computation: similarities compute for the selected pair based on mapping method.

4-similarity aggregation: all similarity metrics are aggregated to produce single one [11].

Ontology alignment is to find the correspondences entities that are equivalent or subsumed relation, and extracting overlapping information over two ontologies. Some alignment methods use one similarity method [12,13], or combination of them to enhance the precision and recall of the alignment methods [14], results of similarities are aggregated dynamically related to absence or presence of features is used [15],also genetic algorithm is used to weights[16],Linear weighted combination(LWC) is used to aggregate similarities[17].

The measures for matching similarity computation can be classified into terminological measures, semantic measures and Structural Measures. Terminology measures are based on surface similarities. The main idea in using such measures is the fact that it happens that usually similar entities have similar names and descriptions across different ontologies. Structural Measures try to realize similarities by considering the relationship of the entities based on the entities in different ontologies are similar when their adjacent entities are similar and structures in the ontology graphs.

This paper proposes a modified method to calculate a structural-based alignment method based on information content taken into account various aspects of the structure of ontologies to recognize related entities. It proposes modifying information content calculation using intrinsic information from the ontology taking into account concept attributes, we determine the similarity by several computing similarity methods for each pairs of concepts.

IC for each concept is calculated by using their attributes(properties, sub-concepts, instance).determine LCS(Least Common Subsumed) for each pair using the pervious calculated similarity , and determine the final similarity value using their relative concepts similarities.Experimental results show that our method performs well comparing with other similar approaches.

The rest of the paper is organized as follows. Section 2 reviews strategies proposed by related works to enable the similarity across different ontologies, while Section 3 presents our approach. Section 4 discusses the results in comparison with related works. The final

Section contains the conclusions and some lines of future research.

II. RELATED WORK

Ontology matching techniques are classified into schema-based and instance based techniques, the schema the similarity between concepts is based on the structure level with ignoring actual data, while instance based similarity is based on data instance of concepts with ignoring structure .There are another classification based on matching techniques to determine the similarities between entities, semantic similarity, terminological similarity and structure similarity [10][18]. Figure 2 indicates to the classifications of matching approaches. The upper classification is based on granularity and input interpretation, the lower classification is based on the kind of input. The middle layer features classes of basic techniques.

Terminological similarity classify to (string-based, language based, and linguistic resources)[19].Structural similarity is viewed ontologies as graph(based on relations between concepts(properties)) and taxonomy(is-a relation) structures containing terms and their inter relationships. Structural similarity can be classified into internal structural and external structural [20, 21]. Also structural similarity based on the shared information between compared concepts (subsumed concept) [22][23].

Semantic similarity measures can be classified into Structure based measures, where it based on the hierarchy structure of ontology based on path length where the shortest path between compared concepts is more similar.

Also by the depth of compared concepts based on the assumption that concepts lower down in the hierarchy are less differentiated than those higher up [24],modifying the pervious similarity by taking into account the number of links between concepts with respect to the max depth of ontology[25]. And Information Content Based Measures(IC), described in the next part in detail.

Alignment methods are used in matching one similarity method or combination of them to enhance the precision and recall of alignment method [26].RIOMM proposes a muti-strategies ontology mapping to reduce risk for selecting these methods [27].

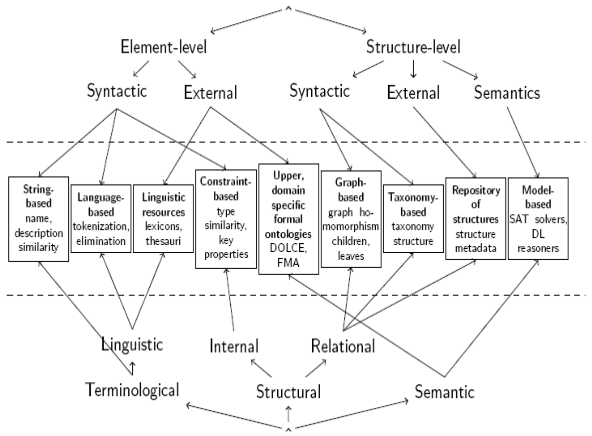


Fig. 2. Classification of matching techniques

A new terminology (lexical) similarity is provided by dealing with each concept as a bag of words and structure similarity by creating neighbor matrix for ontology [28]. Semantic matching (s_match) is a matching method based on linguistic and lexical using wordnet to determine the equivalence and more general and less general relationship between concepts (nodes) [29]. GLUE is a method used multiple learners and exploits information in concept instances where it based on three steps, learn the joint probability distribution of instances of classes of input ontologies, then estimate the similarity between classes based on instances and then filter of the matches result[30].LOM (Lexicon-based Ontology Mapping) finds the morphism between vocabularies in order to reduce human labor in ontology mapping using four methods: whole term, word constituent, synset, and type matching. LOM does not guarantee accuracy or correctness in mappings [31].

Background knowledge is considered as an important approach used in ontology comparing as a bridge to improve the result of matching [32]. Semantic Flow Network (SFN) presents a way to solve compound mapping [33] . Rough set proposed a way to deal with the uncertainty matching and find the final result of matching [34]. BOAT suggests a way to distinguish between trivial and non-trivial matching, where trivial match obtains from string comparator, based on words can equal string but different meaning. Non-trivial matches mean semantic similarity between terms not identical. It takes into account the structure of entities (subclasses, and super classes) and their description (label, comment) [35].

Structure similarity of two elements in distinct models are relies on a pair of elements are similar if their adjacent elements are similar[36].Structural similarity can be classified into internal structural and external structural, internal structural is calculated by comparing the properties of concepts, while external structural is computed by comparing super concepts, siblings ,and sub concepts [37].using path comparator, which is selected by anchor from linguistic similarity, these anchors are determine automatic or semi-automatic by users. it assumes that the paths that connected between two similar terms(anchor) consists of similar entities [38],anchor approach is used as a start point to divide ontologies as segments and match them by string and structure similarities(internal and external structural)[39].LILY uses sub graph to structure similarity , linguistic and semantic similarities[40]. Omen (Ontology Mapping ENhancer) based on Bayesian Net in found matching ,by building network begin with two nodes or concepts that equal string matching and then build network from the structure related concepts to start node which increase the similarity based on the equal concept related to .Then begin Bayesian network for concepts[41]. Structural similarity can be measured based on internal structure of ontology (properties of concepts), external structure (super concepts ,sub concepts and sibling), or by using subsuming concepts (Least Common Subsumed (LCS)) which is obtained from identical equal string similarity , hyponyms comparing , or common features between concepts[42]. Structural similarity is used in semantic similarity where it takes into account the path between concepts and depth between them.

IC of concepts is calculated sometimes with ontology structure, others using wordnet as information source[43-46].Information content have many developed works, start from calculating the IC by negative log of probability of number of occurrences the concept in corpus to total corpus[43].

$$IC(c) = -\log(p(c)) \qquad (1)$$

Where, p(c) is the frequency of concept c in corpus. WordNet is organized in a meaning and principled way. WordNet used in calculation of IC, where the concepts with many hyponyms provides less information, expresses the IC value of a WordNet concept as a function of the hyponyms it has [44,47]. Formally

$$IC(c) = 1 - \frac{\log(hyp(c)+1)}{\log(max\_nodes)} \qquad (2)$$

Where the function $hyp(c)$ returns the number of hyponyms of a given concept and max_nodes is a constant that is set to the maximum number of concepts that exist in the ontology. This method have drawback where concepts with the same sub concepts have the same information content although in different depths, [45] overcome the drawback by complement hyponym-based IC computation with the relative depth of each concept in the ontology. [42] Considers number of leaves of concept compared to the number of taxonomic subsume.

Resink proposed estimating semantic commonalties among concepts based on the amount of information they

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 8, August 2013

share. In taxonomy this information is represented by the least common subsume of both concepts. Increasing IC value means increasing of the semantic similarity between concepts [43].

$$\mathrm{Sim}(c1, c2) = \mathrm{IC}(\mathrm{LCS}(c1,c2)) \qquad (3)$$

The problem of resink similarity is any pair of concepts with the same LCS will result the same similarity exact value. Lin similarity considers the common information content for concepts, and information content of compared concepts [46].

$$\mathrm{sim}(c1, c2) = \frac{2 * \mathrm{IC}\big(\mathrm{LCS}(c1, c2)\big)}{\mathrm{IC}(c1) + \mathrm{IC}(c2)} \qquad (4)$$

[48] Also considers the principle of the common subsume but calculate the similarity but the difference between IC of each concept and the IC of their

LCS.$\mathrm{Sim}(c1, c2) = \mathrm{IC}(c1)+\mathrm{IC}(c2)-2*\mathrm{IC}(\mathrm{LCS}(c1,c2))$
(5)

Information content is used to determine the common concepts, [49] presents the common information content by find the common features of the compared entity classes. A virtual root called as "Anything" was used to connect the considered ontologies.

Matching similarity based on linguistic is considered as analyzing entities in isolation where it is ignoring the relationships with other entities. Similarity taking into account the context of entities is an important similarity for ontology mapping that is achieved by structural similarity. Heterogeneity between ontologies occur when one ontology have more details than other. So this paper solves this problem by structural similarity that considers entities that have common parent have a relationship.

III. THE PROPOSED RELATIVE STRUCTURE SIMILARITY METHOD

The content information approaches presented by Resink and others proposed are based on Lowest Common Sub-sumer (LCS) for single ontology ,they did not consider the importance of concepts (information content) that aligned using the description of concept relative to all data ontology[43, 46,47,50,51] .

This paper proposes a structure similarity method based on relative concepts similarity and information content, modifying information content by taking into account concept attributes which also indicates the importance of classes(more information attributes describe class is increasing the importance class) . The method determines equivalence, subsume relationship depending on the Owl ontology features. Hence more information description for ontology entities (concept, properties, instances, equivalent, disjoint, union, complement,…) are presented .The alignment method architecture used in this paper is shown in figure (3), the architecture consists of three layers preprocessing, primitive similarity, and relative structure similarity. The alignment process is to determine the correspondence between two input ontologies entities. It includes three steps:

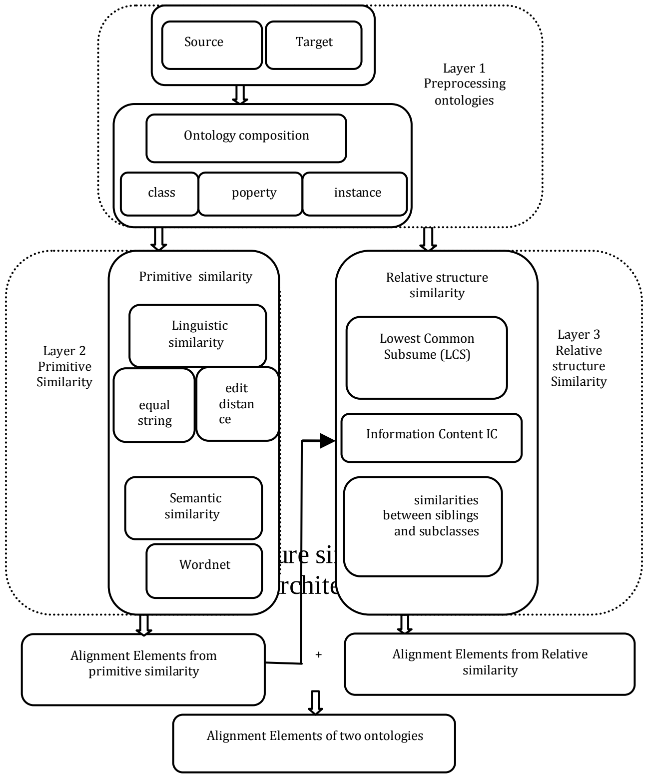


Fig. 4. A fragment of owl ontology.

A. *ONTOLOGY Pre-PROCESSING*

In this paper ontologies are described within the knowledge representation language OWL.

OWL is a computer language used to write ontologies which provide more vocabularies for describing objects), the preprocessing of owl ontology is used to. The pre-processing layer has two phases, the first phase picks up all information about ontology concepts, properties (object property-data type property) and instances, while the second phase analyzes the vocabularies content by splitting(tokens) and stemming, where it consist of tokens "name chunk of several tokens" such as punctuation, upper cases characters, symbols and special characters, for example "SystemsStaff" is splitted into "system" and "staff", also abbreviation and acronyms are enlarged. Such as token « SW » which referred to "Semantic Web" this expansion is performed using an external dictionary.

The tokens of the two ontologies are then converted to lowercase characters for the comparison .The same thing is done for concepts comments.

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 8, August 2013

Figure 4 shows a part of owl ontology describing an organization, in first layer it picks up data with gray highlight and stemming them.

B. PRIMITIVE SIMILARITY

In this paper, ontologies are described by OWL, an entity in ontology is defined as: $e \in C \cup P$ where C and P are the sets of concepts and properties in ontology respectively.

We first compute the initial similarities between entities and then used this initial to select LCS from two ontologies for each pair of concepts. When calculating similarities between entities, we aim to maximize the descriptive (or semantic) information of an entity, such as its ID, its label and its comment to cover diverse situations. The descriptive information of an entity composed of concept descriptive and property descriptive, concept descriptive (label, comment), and property descriptive (domain, range, and property label).

In this layer primitive similarities are calculated using linguistic similarity and semantic similarity for entities descriptive. After converting each concept Id, label, and comment to tokens, these tokens are used in the comparison in the case of terminology similarity equal string and using edit distance similarity. Edit distance estimates the number of operations needed to convert one string into another, the similarity of two labels of concepts e1 and e2 which defined as

$$Ed(e1, e2) = 1 - \#op / (max\_length(l(e1), l(e2)))$$

Where,

#op: indicates the number of operations.

max_length(l(e1), l(e2)) :represents the maximal length of the two labels.

$$sime(e1,e2) = \begin{cases} 1 & \text{if } e1 = e2 \\ 0 & \text{if } e1 \neq e2 \end{cases}$$

Also comment of concept is converting to VS (vector similarity) and calculates the similarity for vectors taking into account the frequency of tokens. The total similarity between two concepts is calculated as the summation of all results similarities.

WordNet is an electronic lexical database developed at Princeton University. Wordnet entries ("senses") are organized into synonyms sets ("synsets") representing concepts. Each synset (synonyms set) in WordNet is followed by its definition ("gloss") which contains a defining phrase, an optional comment and examples. WordNet supports two types of relations: semantic relations, which link concepts (i.e. synsets), such as hypernymy, hyponymy, meronymy, holonymy, etc. and lexical relations, such as antonymy, which links individual words [52].

WordNet is used to determine the semantic similarity and subsume relationship between concepts, by obtaining the synonyms of each concepts id, and label to compare them. Wordnet and ontology are used to determine the subsume relationships (more than (hypernyms), less than


```
<owl:Class rdf:ID="Institution">
<rdfs:subClassOf
rdf:resource="http://xmlns.com/foaf/0.1/Organization"/>
<rdfs:label xml:lang="en">Institution</rdfs:label>
<rdfs:comment xml:lang="en">An institution.</rdfs:comment>
<rdfs:subClassOf>
<owl:Restriction>
<owl:onProperty rdf:resource="#name"/>
<owl:cardinality
rdf:datatype="http://www.w3.org/2001/XMLSchema#nonNegati
veInteger">1</owl:cardinality>
</owl:Restriction>
</rdfs:subClassOf>
<rdfs:subClassOf>
…
<owl:onProperty rdf:resource="#address"/>
<owl:maxCardinality
rdf:datatype="http://www.w3.org/2001/XMLSchema#nonNegati
veInteger">1</owl:maxCardinality>
</owl:Restriction>
</rdfs:subClassOf>
</owl:Class>
……
<owl:ObjectProperty rdf:ID="institution">
<rdfs:domain rdf:resource="#Report"/>
<rdfs:range rdf:resource="#Institution"/>
<rdfs:label xml:lang="en">institution</rdfs:label>
<rdfs:comment xml:lang="en">The sponsoring institution of a
technical report.</rdfs:comment>
</owl:ObjectProperty>
……
−<owl:Class rdf:ID="School">
<rdfs:subClassOf rdf:resource="#Institution"/>
<rdfs:label xml:lang="en">School</rdfs:label>
<rdfs:comment xml:lang="en">A school or
university.</rdfs:comment>
</owl:Class></owl:Class>
……
−<owl:ObjectProperty rdf:ID="school">
<rdfs:range rdf:resource="#School"/>
<rdfs:label xml:lang="en">school</rdfs:label>
<rdfs:comment xml:lang="en">The name of the school where a
thesis was  written.</rdfs:comment>
</owl:ObjectProperty>
```


(hypnoym)).

$$sims(e1,e2) = sim(gloss(e1),gloss(e2)) + sim(hypo(e1),hypo(e2)) + sim(hyper(e1),hyper(e2)) + sim(synst(e1),synst(e2))$$

The same way is used for properties similarity. The flow chart representing primitive similarity is shown in figure (5).

Figure 5 represent the flow chart for primitive similarity, where we read two ontologies o1,o2 and wordnet database, and collect all information from two ontologies, c1[n1],c2[n2] represent concepts where n1,n2 represent total number of concepts in o1,o2 respectively ,also properties of each concept represent in vector p[np1],p[np2],where np1,np2 are number of properties for each concept. Then we process concepts by splits or

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 8, August 2013

expand each concept and property, for each token we calculate the linguistic similarity (equal string (sime), edit distance(Ed)), semantic similarity (sims), finally we store all the calculated similarity values.

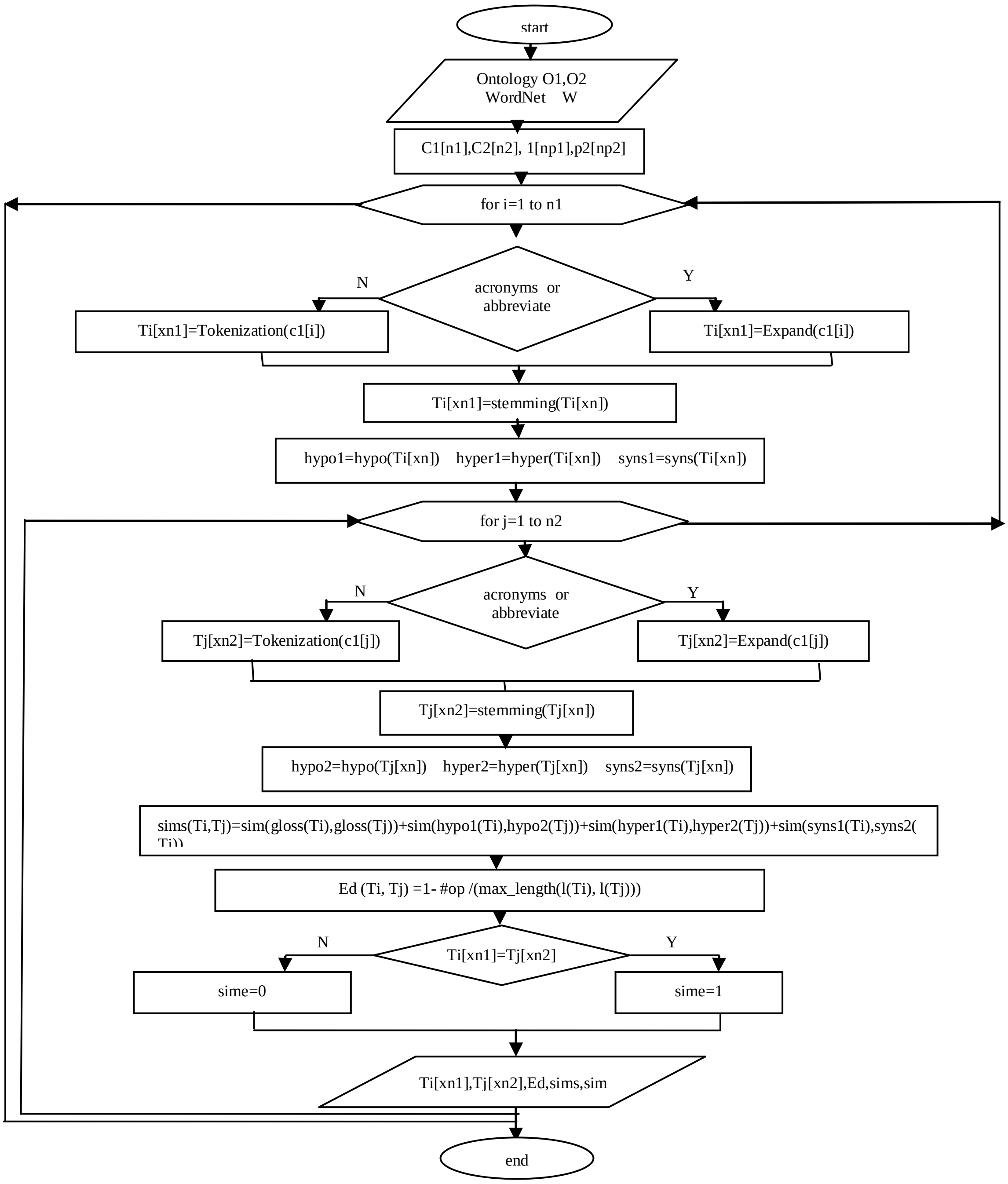


Fig. 5. Flow chart for primitive similarity

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 8, August 2013

RELATIVE STRUCTURE SIMILARITY

The relative similarity is based on structure similarity and information content. This similarity depends on the common structure (super class) plus their information content (concept attributes) for the concepts. Resink basically works on a single ontology by calculating the similarity between two classes in the same ontology, so there is no problem to search for the common super class. A modified version of Resink defines a virtual root to connect the two ontologies and considers that root as a common super concept, or deal with common concepts that class have common features between two concepts[49].

This paper provides a method to calculate the structure similarity using relative similarity and information content of the concepts, calculating relative similarity based on the structure of concepts, so we need to determine LCS in two ontologies, by estimating the similarity of two super concepts resulted from the second layer. The information content is calculated from the importance class that take into account all the class attributes, finally the similarities between each two class are calculated from their relative classes and their information content. A flow chart of the proposed method is shown in figure (6).

Figure 6 represent a flow chart that describe the relative similarity, start by taking two concepts from o1,o2 stored in c1[n1],c2[n2],obtain a list of super classes for each concept sup1[xn1],sup2[xn2], where xn1.xn2 are number of super concepts for each concept in o1,o2.

For two concepts calculate the information content (IC) taking into account number of properties (prop (c1)), instances (inst(c)), and sub classes (sub(c)). For each concept in sup1, sup2 we obtain the similarity values stored if there is a similar value suppose it as LCS until finish all array, if no similar pair we read another pair from concepts. For each concept obtain their sub concepts sub1[xn1],sub2[xn2] ,compare each pair to obtain similarity and information content IC(sub1[i],sub2[j]) taking into account IC(lcs) to calculate the total similarities for them sim(sub1[i],sub2[j]).

Owl ontology is a formal and explicit conceptualization of description. Every concept is defined by property functions, sub concepts, instances and constraints. Although domain ontology are not completed as wordnet in terms of concepts, so a sufficient IC of concept without using external large text corpora can be generated using intrinsic information of the concept. Concept attributes are used as intrinsic information.

The pervious calculation of IC in ontology is obviously concepts and concept hierarchy. However, OWL ontology also contains properties, restrictions and relations. Properties are used to define functionality of a concept explicitly to specify a meaning. They are related to concept by means of domain, range and restrictions. Also it contains instances where concepts plays important role in classification of instances which act as knowledge for this concept.

To improve the information content of the concept. All attributes that describe the concept is taken into account with respect to total ontology attributes. The information content is measured by concept relations (properties) ICp(c), concept instances ICi(c) and their sub concepts ICc(c). The information content of concept is calculated by negative logarithm of the probability of concept in text or any information used to describe concept. Where it is inversely related with attributes of some parameters and directly proportional with other attributes.

$$\mathrm{ICc}(c) = 1 - \frac{\mathrm{LOG}(\mathrm{sub}(c)+1)}{\mathrm{LOG}(\mathrm{t\_class})} \qquad (6)$$

$$\mathrm{ICp}(c) = \frac{\mathrm{LOG}(\mathrm{prop}(c)+1)}{\mathrm{LOG}(\mathrm{t\_prop}+1)} \qquad (7)$$

$$\mathrm{ICi}(c) = \frac{\mathrm{LOG}(\mathrm{inst}(c)+1)}{\mathrm{LOG}(\mathrm{t\_inst}+1)} \qquad (8)$$

where ICc(c) is the information content based on the sub concepts of the class c, more sub concepts of the class (sub(c)) lead to less information it express [44], and t_class is the constant value for ontology that is the total number of ontology classes.

ICp(c) estimate the information content based on relations (properties),where prop(c) denotes to the number of properties of concept c, where information content is negative log of information(external source or intrinsic information) adding one in log variable to remove log zero value , and t_prop represents the total number of properties available in ontology.

ICi(c) is the information content based on instances of class c(inst(c)) and t_inst represents the total instances available in ontology.

The total information content is defined as the sum of all ICs (property information content, instance information content and subclasses information content) weighted to total ontology attributes, as shown in equation 9.

$$\mathrm{IC}(c) = \mathrm{w1} * \mathrm{ICp}(c) + \mathrm{w2} * \mathrm{ICi}(c) + \mathrm{w3} * \mathrm{ICc}(c) \qquad (9)$$

Where,

$$\mathrm{w1} = \frac{\log(\mathrm{t\_prop}+1)}{\log(\mathrm{t\_prop}+1) + \log(\mathrm{t\_class}) + \log(\mathrm{t\_inst}+1)} \qquad (10)$$

$$\mathrm{w2} = \frac{\log(\mathrm{t\_inst}+1)}{\log(\mathrm{t\_prop}+1) + \log(\mathrm{t\_class}) + \log(\mathrm{t\_inst}+1)} \qquad (11)$$

$$\mathrm{w3} = 1 - (\mathrm{w1} + \mathrm{w2}) \qquad (12)$$

w1, w2, w3 are weights for the property information, instance information and subclasses information respective which are defined by natural of an ontology, where it increase by increasing the number of relations (properties t_prop) and number of instances(t_inst) in ontology, on the contrary while a small number of attributes with a large number of ontology concepts(t_class) lead to decrease correspondence weight

$$\mathrm{IC}(c1,c2) = |\mathrm{IC}(c1) * \mathrm{IC}(c2)| \qquad (13)$$

$$\mathrm{sim}(c1,c2) = \big(\mathrm{ws} * \mathrm{Sims}(c1,c2) + \mathrm{wl} * \mathrm{Siml}(c1,c2)\big)\mathrm{IC}(c1,c2) \qquad (14)$$

The information content (importance) of alignment concepts c1, c2 represents by (11) which is used in the

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 8, August 2013

initial similarity between compared classes (12). $\mathrm{Sims}(c1, c2)$ is the semantic similarity between two concepts and $\mathrm{Siml}(c1, c2)$ is the linguistic similarity between the same concepts.

$$\mathrm{ICr}(c1, c2) = \frac{2 * \mathrm{ICc}(c1, c2)}{\mathrm{IC}(c1) + \mathrm{IC}(c2)} \quad (15)$$

$$\mathrm{ICc}(c1, c2) = \max(\mathrm{Ics}(c1), \mathrm{Ics}(c2)) \quad (16)$$

$$\mathrm{simr}(c1, c2) = \left(\mathrm{ws} * \mathrm{Simrs}(c1, c2) + \mathrm{wl} * \mathrm{Simrl}(c1, c2)\right) * \mathrm{ICr}(c1, c2) \quad (17)$$

Relative similarity between two classes is relative to their sibling, super classes, subclasses similarities .So to calculate the similarity between the correspondence classes from two ontologies using semantic similarity and linguistic similarity, taking into account the information content that calculated using attributes indicates to importance of classes.

ICr(c1,c2) is the information content for relative classes which based on the Least Common Subsume concept(LCS) ICc(c1,c2) in (15).

Sims(c1,c2) , Simrs (c1,c2): semantic similarity between two classes c1,c2, and their relative respectively.

Siml(c1,c2), Simrl (c1.c2) : linguistic similarity between two classes c1,c2, and their relative respectively.ws ,wl :two weighted value for semantic similarity and linguistic similarity.

$$\mathrm{Nsim}(c1, c2) = \mathrm{sim}(c1, c2) + \sum_{i=1}^{k} \mathrm{simr}(c1i, c2i) \quad (18)$$

IV. EVALUATION AND RESULTS

A. Evaluation

The outputs of the alignment are a set of pairs that have relationships, Measurement of performance and comparison of methods of alignment are the main ways to estimate the alignment performance, as well as the evaluation of the quality of alignment. This evaluation is made in two steps the first is manually solving the alignment by finding the correspondence between ontologies, which consider as alignment of reference, then comparing the alignment method with the reference. The results are three values Nfound, Nexpected and Ncorrect.

$$\mathrm{Precision} = \frac{|\mathrm{Ncorrect}|}{|\mathrm{Nfound}|}$$

$$\mathrm{Recall} = \frac{|\mathrm{Ncorrect}|}{|\mathrm{Nexpected}|}$$

$$\mathrm{F-measure} = \frac{2 * \mathrm{pecision} * \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}$$

Where, Nfound represent the output of alignment method, The Nexpected is the pair result of reference alignment. The Ncorrect is the intersection of the two values of alignment method result and the reference alignment result [53].

Precision measures the fraction of found alignments that are actually correct. When the precision equal to 1 that mean all reference pairs are found in the alignment results ,but it does not mean that all alignments result are in references pairs. Recall measure the fraction of correct alignment to the total number of correct existing alignment. A high recall indicates too many of alignments have actually been found. If there is a high recall and low precision this indicate to many false alignments-measure represents the harmonic mean of precision and recall. This will be the main measure of quality assessment. Due to the impact of the budget, it makes sense to draw the precision and recall against each other. Thus, one sees that the accuracy and / or call the method that works best [10]. Alignment is the process of founding the correspondences between concepts of two ontologies or other relationships between them, sometimes these concepts is not important concept in ontology mean concepts not provide detail information in the description of domain, where centrality of concept and its density increase the weight of concept importance. The importance of concept is the measurement it take into account the concept's attributes and other related concepts.

B. Datasets

We have been running a subset of the OntoFarm dataset for the matching systems participating in OAEL. All ontologies model the same domain is conference organization, based on different conferences. Ontologies reflect different conceptualizations of the same domain, this way simulating 'real-world heterogeneity' of semantic web ontologies, using complete reference alignment for dataset can be downloaded from http://nb.vse.cz/~svabo/oaei2010/. Table I present the ontologies used in our evaluation indicating number of concepts and properties in each ontology.

TABLE I

Oontology	Concepts_no.	Property_no.
CMT.owl	29	59
conference	59	64
confof	38	36
edas	103	50
Ekaw	73	33
Iasted	140	41
sigkdd	49	28

A dataset used in the experiment and their sizes

C. Results

The Relative similarity method has been implemented using the Java programming language with single thread. We can evaluate the new alignment method using the three measures of precision, recall and f-measure. Table II presents the average values of precision ,recall and f-measure for all methods indicating the improvement of the new method (RSS). Fig. 5. represent the result compared with Boat, Flood anchor, s_match and Lom [35,39,29,31]. From Fig. 7, we can see that Precision of RSS better than all compared similarity methods, but there are three pairs of ontologies have more precision than others. Recall is better than all other alignment compared method.

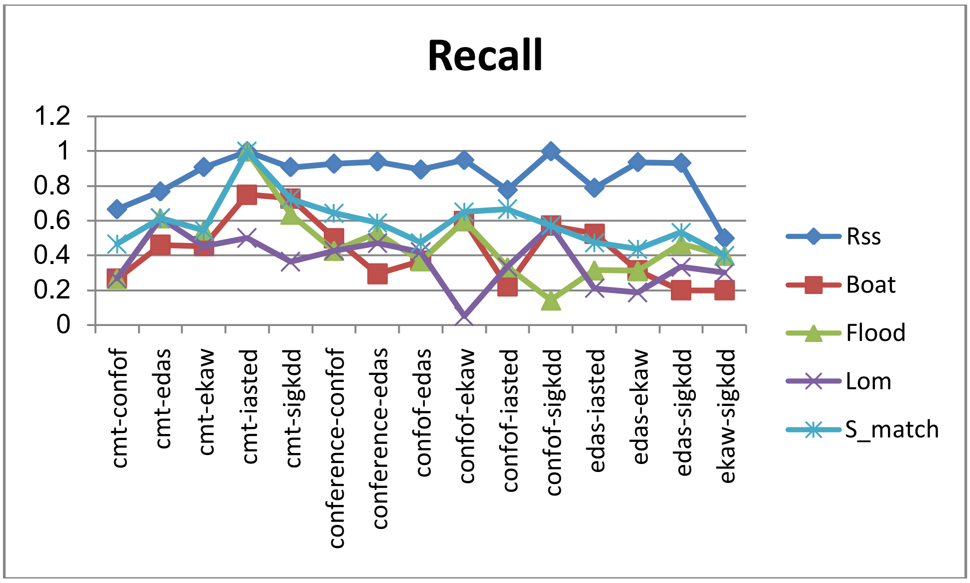
ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 8, August 2013

TABLE II
The average Precision, recall and F-measure of OAEI ontologies by alignment methods

	Recall	Precision	F_measure
Rss	0.860443	0.331012	0.463957
S_match	0.586145	0.164238	0.249786
Flood	0.464095	0.118583	0.180415
Boat	0.430335	0.114663	0.16709
Lom	0.367104	0.187055	0.230306

This method presents a way to estimate the similarity of classes, it is consider as measuring similarity by taking into account the structure similarity plus class importance. This method enhance the precision alignment and recall in some cases, where it increase the number of classes aligned .it also provides a way to align ontologies that have different languages in the same ontology, where the string comparator and WorldNet which serve English language only are not effective, but here the structure similarity that take into account the context of entities will affect the total similarity of entity.



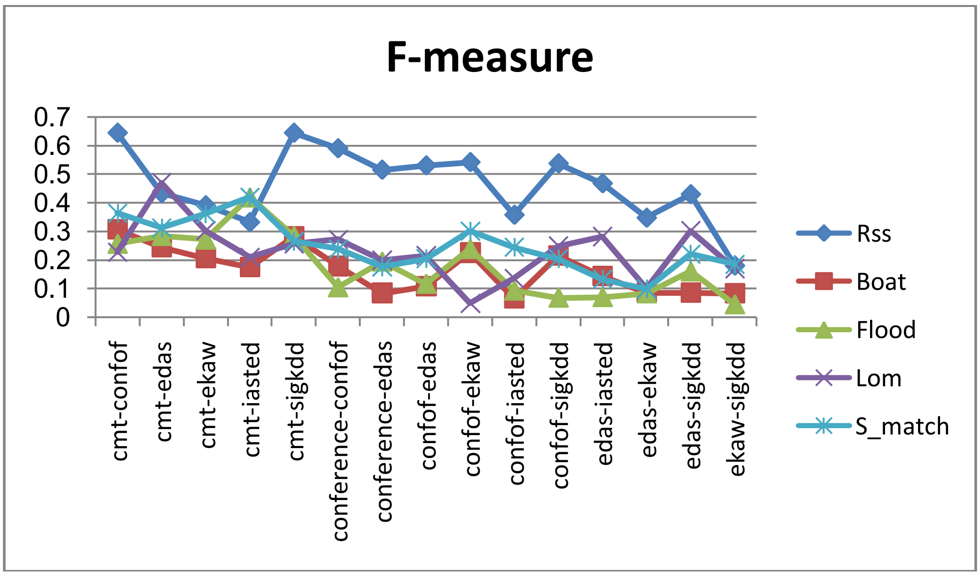


Fig. 7. Comparison of Precision, Recall and F-measure between RSS, BOAT, Flood ,S-match and Lom on dataset ontologies.

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 8, August 2013

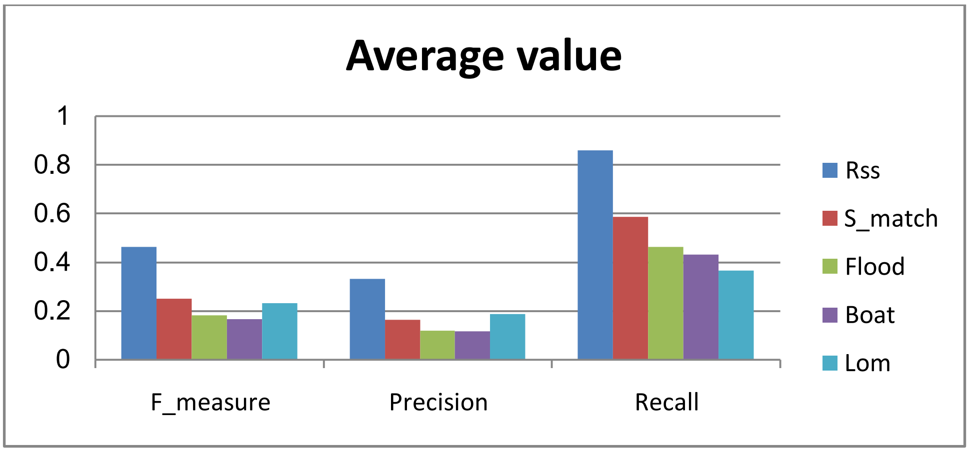


Fig.8. Average values of

CONCLUSION

Ontologies are backbone of the semantic web and many applications. Application deal with different ontologies for the same domain differs in their structure or modeling. Ontology mapping is an important step to deal with different ontologies to compute the relationships between concepts or to estimate the most accurate knowledge when the concept overlap with in multiple ontologies.

This paper presents a method for ontology alignment by using structure similarity for ontologies entities relative to the string similarity and semantic similarity of related entities to concepts compared, taking into account their information content with a modification of IC calculation.

Experimental result showed the superiority of the proposed method over others.

In the future work we can determine how to determine uncertain alignment and how to select the alignment if the concept has many similarity values to many classes.

REFERENCES

[1] N. Shadbolt , W. Hall and T. Berners-Lee," The semantic web revisited," IEEE Intelligent Systems 21 ,2006 .

[2] C. Badica, G. Mangioni and S. Rahimi,"Intelligent distributed information systems," Information Sciences – Special Issue on Intelligent Distributed Information Systems ,2010.

[3] K. Hussein Shafa'amri and J. Omer Atoum ,"A Framework for Improving the Performance of Ontology Matching Techniques in Semantic Web," (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 3, No. 1, 2012.

[4] Sridevi.U.K and N. Nagaveni,"Ontology based Similarity Measure in Document Ranking," International Journal of Computer Applications (0975 - 8887) Volume 1 – No. 26, 2010.

[5] J. Euzenat, and I. Rhône-Alpes & LIG," Introduction to ontology matching and alignment," 2007.

[6] M. Ehrig and Y. Sure," Ontology mapping - an integrated approach," In Proceedings of the European Semantic Web Symposium (ESWS), pages 76–91, 2004.

[7] M. Nagy and M. Vargas-Vera,"Multi agent Ontology Mapping Framework for the Semantic Web,"IEEE transactions on systems, man and cybernetics—part a: systems and humans, vol. 41, no. 4, July 2011.

[8] L.Laera , V. Tamma , T.J.M. Bench-Capon and J´erˆome Euzenat,"Agent-based Argumentation for Ontology Alignments," In Proceedings of the Workshop on Computational Models of Natural Argument (CMNA), 2006.

[9] J. J. Jung ,"Exploiting multi-agent platform for indirect alignment between multilingual ontologies: A case study on tourism business," Expert Systems with Applications, Volume 38, p. 5774-5780 issue 5 May, 2011.

[10] Y. Lv and Ch. Xie," A Framework for Ontology Integration and Evaluation,"Processing third international conference on intelligent networks and intelligent systems, pages 521-524,2010.

[11] M. Ehrig ,S. Staab and Y. Sure, "Bootstrapping Ontology Alignment Methods with APFEL," Processing www special interest tracks and posters of the 14th international conference on World Wide Web, 2005.

[12] D. Engmann and S. Massmann,"Instance Matching with COMA++," In: Proc. of the Workshop on Daten bank systeme in Business, Technologie und Web (BTW 2007), Aachen, Germany 2007.

[13] D. Aumueller, H.-Hai Do, S. Massmann and E. Rahm,"Schema and Ontology Matching with COMA++,"Processing the 2005 SIGMOD international Conference on management of data, p. 906-908,2005.

[14] R. Tous and J. Delgado," A Vector Space Model for Semantic Similarity Calculation and OWL Ontology Alignment," DEXA 2006, LNCS 4080, pp. 307–316, 2006.

[15] Y. R. Jean-Mary and M. R. Kabuka,"ASMOV: Ontology Alignment with Semantic Validation,"September 24, 2007.

[16] A.H. Haddad and A. Selah," Ontology Alignment with FOAM++," International Journal of Computer Applications (IJCA) (0975 – 8887) Volume 18– No.8, March 2011.

[17] I. F. Cruz, F. P. Antonelli, and C. Stroe,"Efficient Selection of Mappings and Automatic Quality-driven Combination of Matching Methods," Proceedings of the 4th International Workshop on Ontology Matching, Vol. 551 http://ceur-ws.org: CEUR Workshop Proceedings ,October 2009.

[18] A. Kh. Alasoud,"Multi-Matching Technique for Combining Similarity Measures in Ontology Integration,"A Thesis In the department of Computer Science and Software Engineering, February 2009.

[19] F. Giunchiglia and P. Shaiko,"Semantic matching,"The knowledge engineering review 18, (03) 265–280, 2004.

[20] S. Jan, I.A. Shah, I. Khan, F. Khan and M. Usman,"Similarity Measures and their Aggregation in Ontology Matching," International Journal of Computer Science and Telecommunications ,Volume 3, Issue 5, May 2012.

[21] G. Stoilos, G. Stamou and S. Kollias,"A string metric for ontology alignment," Proceedings of the 4th International Semantic Web Conference (ISWC2005), pp. 623 - 637,2005.

[22] K.Saruladha, Dr.G.Aghila, and A.Bhuvaneswary,"Computation of semantic similarity among cross ontological concepts for biomedical domain," Journal of computing, volume 2, issue 8, august 2010, issn 2151-9617,2010.

[23] D. Sánchez, A. Solé-Ribalta , M. Batet and F. Serratosa,"Enabling semantic similarity estimation across multiple ontologies: an evaluation in the biomedical domain," J Biomed Inform,2012.

[24] Z. Wu and M. Palmer,"Verb semantics and lexical selection,"In: 32nd Annual meeting of the association for computational linguistics. Las Cruces (New Mexico): Association for Computational Linguistics, p. 133–38, 1994.

[25] C. Leacock and M. Chodorow ,"Combining local context and WordNet similarity for word sense identification. In WordNet: an electronic lexical database," MIT Press, p. 265–83,1988.

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 8, August 2013

[26] M. Houshmand and E. Khorram,"Similarity Aggregation in Ontology Matching Based on Reliability Maximization,"Electrical Engineering (ICEE), 19th Iranian Conference ,17-19 May 2011.
[27] J.Li, J. Tang, Y. Li and Q. Luo,"RiMOM: A Dynamic Multistrategy Ontology Alignment Framework," IEEE transactions on knowledge and data engineering, vol. 21, 2009.
[28] I. Akbari and M. Fathian,"A novel algorithm for ontology matching,".Journal of Information Science, Vol.36, pp. 324–334,2010.
[29] F. Giunchiglia, P. Shvaiko and M. Yatskevich ,"S-MATCH: AN ALGORITHM AND AN IMPLEMENTATION OF SEMANTIC MATCHING," In proceedings of the European Semantic Web Symposium, LNCS3053,pp. 61-75, 2004.
[30] N. Choi, Il-Y. Song and H. Han,"A Survey on Ontology Mapping," ACM sigmod record, Volume 35 Issue 3,pages 34 – 41, September 2006.
[31] J. Li,"LOM: A Lexicon-based Ontology Mapping Tool," Proceedings of the Performance Metrics for Intelligent Systems (PerMIS. '04), 2004.
[32] V. Mascardi, A. Locoro, and P. Rosso,"Automatic Ontology Matching via Upper Ontologies: A Systematic Evaluation," IEEE Transactions on knowledge and data engineering, Vol. 22, No. 5, may 2010.
[33] V. Fionda and G. Pirr´o,"Semantic Flow Networks: Semantic Interoperability in Networks of Ontologies," Proceedings of the 1st Joint International Semantic Technology Conference (JIST 2011).
[34] S. Jan, M. Li1, Gh. Al-Sultany and H. Al-Raweshidy," Ontology Alignment using Rough Sets. Fuzzy Systems and Knowledge Discovery (FSKD)," 8th International Conference on, 26-28 July 2011.
[35] W. W.Kh. Chua and J.-j. Kim,"BOAT: Automatic alignment of biomedical ontologies using term informativeness and candidate selection," Journal of Biomedical Informatics (Elsevier), Volume 45, Issue 2, Pages 337-349, April 2012.
[36]S. Melnik, H. Garcia-Molina and E. Rahm," Similarity flooding: a versatile graph matching algorithm and its application to schema matching," Proceedings of the 18th International conference on Data Engineering (ICDE 2002), San Jose,CA, USA, pp. 117 – 128, 2002.
[37]G. Stoilos, G. Stamou and S. Kollias,"A string metric for ontology alignment," Proceedings of the 4th International Semantic Web Conference (ISWC2005), pp. 623 – 637, 2005.
[38] N. F. Noy and M. A. Musen,"Anchor-PROMPT: Using Non-Local Context for Semantic Matching," In Proc. workshop on ontology and information sharing IJCAI,2001.
[39] Md. S. Hanif, and M. Aono,"An efficient and scalable algorithm for segmented alignment of ontologies of arbitrary size,"Journal Web Semantics: Science, Services and Agents on the World Wide Web, Volume 7, Pages 344-356, Issue 4, December 2009.
[40] P. Wang and B. Xu,"Lily: Ontology Alignment Results for OAEI 2009," In the 4th International Workshop on Ontology Matching, Washington Dc., 2009.
[41] P. Mitra, N. F. Noy and A. R. Jaiswal,"OMEN: A Probabilistic Ontology Mapping Tool," In Workshop on Meaning Coordination and Negotiation at the Third International Conference on the Semantic Web ,ISWC,2004.
[42] D.Sanchez and M. Batet,"Semantic similarity estimation in the biomedical domain:An ontology-based information-theoretic perspective," Journal of Biomedical Informatics, Vol. 44, Issue 5, Pages 749-759, October 2011.
[43] P.Resnik, "Using information content to evaluate semantic similarity in taxnonomy," IJCAI'95 Proceeding of the 14th international joint conference on Artificial intelligence-Vol.1.p448-453, 1995.
[44] N. Seco, T. Veale and J. Hayes,"An intrinsic information content metric for semantic similarity in WorldNet," Proceedings of ECAI'2004, the 16th European Conference on Artificial Intelligence Valencia, Spain, 2004.
[45]Zhou Z, Wang Y and Gu J. ,"A new model of information content for semantic similarity in WorldNet," FGCNS '08. Second International Conference, Future Generation Communication and Networking Symposia, IEEE computer Society, 2008.
[46] D. Lin ,"An information-theoretic definition of similarity," Proc. 15th International Conference on Machine Learning , pp. 296-304,1998.
[47] G. Pirro, "A semantic similarity metric combining features and intrinsic information content," Journal of Data and Knowledge Engineering 68 , 1289–1308,2009.
[48] J. J. Jiang and D.W. Conrath,"Semantic similarity based on corpus statistics and lexical taxonomy,"In Proceedings of International Conference Research on Computational Linguistics (ROCLING X), Taiwan, p. 19–33, 1997.
[49]K. Saruladha, Dr.G. Aghila,and A. Bhuvaneswary,"COSS: Cross Ontology Semantic similarity Measure-An Information Content Based Approach,"IEEE-International Conference on Recent Trends in Information Technology, ICRTIT 2011,MIT, Anna University, Chennai. June 3-5, 2011.
[50] J. Jiang and D. Conrath,"Semantic similarity based on corpus statistics and lexical taxonomy," Proceedings of the 10th International Conference on Research in Computational Linguistics ,1997.
[51] D. Sanchez, M. Batet and D. Isern,"Ontology-based information content computation ,"Journal Knowledge-Based Systems, Volume 24, Issue 2, Pages 297-303, March 2011.
[52] J. Michelizzi, S. Banerjee and T. Pedersen,"WordNet::Similarity: measuring the relatedness of concepts," processing Hlt-Naacl-Demonstration,pages 38-41,2004.
[53] A. Souhaib, Kh. Mohamed and El K. K. Eddine, Ontology Alignment OWL-Lite. eLearning - Theories, Design, Software and Applications, Patrizia Ghislandi (Ed.), ISBN: 978-953-51-0475-9, InTech, Available from: http://www.intechopen.com/books/elearning-theories-design-software-and- applications/ontology-alignment-owl-lite,2012.